DOCUMENT RESUME

ED 311 077 TM 013 917

AUTHOR Berger, Martijn P. F.
TITLE On the Efficiency of IRT Models When Applied to Different Sampling Designs. Project Psychometric Aspects of Item Banking No. 45.
INSTITUTION Twente Univ. of Technology, Enschede (Netherlands).
REPORT NO 89-4
PUB DATE Jun 89
NOTE 37p.
AVAILABLE FROM Biblictheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Difficulty Level; *Efficiency; Equations (Mathematics); *Heuristics; *Item Sampling; *Latent Trait Theory; *Mathematical Models; Test Items
IDENTIFIERS One Parameter Model; Parametric Analysis; Three Parameter Model; Two Parameter Model; *Variance (Statistical)

ABSTRACT

The problem of obtaining designs that result in the most precise parameter estimates is encountered in at least two situations where item response theory (IRT) models are used. In so-called two-stage testing procedures, certain designs that match difficulty levels of the test items with the ability of the examinees may be located. Such designs have the advantage that the variance of the estimated parameters can be controlled. In situations in which IRT models are applied to different groups, efficient multiple-matrix sampling designs are applicable. The choice of matrix sampling designs will also influence the variance of the estimated parameters. In this paper, heuristic arguments are given to formulate the efficiency of a design in terms of an asymptotic generalized variance criterion, and a comparison of the efficiencies of several designs is made. It is shown that some designs may be found to be most efficient for the one-parameter model and two-parameter model, but not necessarily so for the three-parameter model. Five figures are included. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED311077

On the Efficiency of IRT Models when Applied to Different Sampling Designs

Research Report 89-4

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. NELISSEN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Martijn P.F. Berger

department of EDUCATION

University of Twente

Division of Educational Measurement and Data Analysis

BEST COPY AVAILABLE

TM013917

Project Psychometric Aspects of Item Banking No.45

Colofon:
Typing: L.A.M. Bosch-Padberg
Cover design: Audiovisuele Sectie TOLAB Toegepaste Onderwijskunde
Printed by: Centrale Reproductie-afdeling

On the Efficiency of IRT Models when Applied to Different Sampling Designs

Martijn P.F. Berger
University of Twente
Enschede, The Netherlands

On the efficiency of IRT models when applied to different sampling designs , Martijn P.F. Berger - Enschede : University of Twente, Department of Education, June, 1989. - 30 pages

Abstract

The problem of obtaining designs that result in the greatest precision of the parameter estimates is encountered in at least two situations where item response theory Models are used. In so-called two-stage testing procedures certain designs that match difficulty levels of the test items with the ability of the examinees may be located. Such designs have the advantage that the variance of the estimated parameters can be controlled. In situations where IRT models are applied to different groups, efficient multiple-matrix sampling designs are applicable. The choice of matrix sampling designs will also influence the variance of the estimated parameters.

In this paper heuristic arguments are given to formulate the efficiency of a design in terms of an asymptotic generalized variance criterion and a comparison of the efficiencies of several designs is made. It is shown that some designs may be found to be most efficient for the one- and two-parameter model, but not necessarily so for the three-parameter model.

Key words: Item Response Theory; Efficiency; Optimal design; Generalized variance

On the Efficiency of IRT Models when Applied to Different Sampling Designs

The notion of information about parameters in item response theory (IRT) models has lead to several applications. For example, in test construction and item selection, Theunissen (1985), van der Linden (1987) and van der Linden and Boekkooi-Timminga (1988) used the information about the ability parameters to obtain optimal item selection procedures and optimal test designs. Samejima (1977) showed how the information as a function of ability could be used in tailored testing. Lord (1974), Lord and Wingersky (1985) and Thissen and Wainer (1982), among others used the asymptotic standard errors, obtained from the inverse of the information about the parameters, to compare the relative efficiency of tests, models and designs. See Lord (1980) and Hambleton and Swaminathan (1985) for applications of information as a function of the IRT parameters in various fields of measurement.

Lord (1962) showed that for a fixed number of item-person confrontations the group mean of an item domain is estimated most reliably, when each item is taken by a different sample of persons, and a similar result was obtained empirically by Pandey and Carlson (1976). Although this result has been used to stress the importance of multiple-matrix sampling designs, the question still remains whether these designs are also efficient to estimate the IRT parameters.

In this paper the efficiency of various designs is considered in terms of a generalized variance criterion connected with IRT parameters and this criterion is used to compare the efficiencies of designs for the one-, two- and three-parameter models. The results of this comparison are relevant to constructors of item banks who need to estimate the item parameters efficiently.

First, however, the use of information in IRT models is discussed and heuristic arguments are given to propose the generalized variance criterion.

Information in IRT Models

In any statistical model the notion of information is connected with the estimation of unknown parameters. The amount of information is defined (Kendall & Stuart, 1973, p. 10) as

$$(2) \qquad Inf = E\left[\frac{\delta \ln(L)}{\delta \theta}\right]^2 ,$$

where L is the likelihood function of a parameter θ and the information about the parameters in IRT models is defined similarly. For example, consider the three-parameter logistic model, which gives the probability of a correct response to item i (i=1,...,n) as a function of the ability parameter θ_a $\in <-\infty, +\infty>$ for examinee a (a=1,...,N):

$$(3) \qquad P_i(\theta_a) = c_i + (1-c_i)\ \{1+\exp[-a_i(\theta_a-b_i)]\}^{-1}$$

where $b_i \in \langle -\infty, +\infty \rangle$ and $a_i \in \langle 0, +\infty \rangle$ are the item difficulty and discrimination parameters, respectively, and $c_i \in \langle 0,1 \rangle$ is the guessing parameter. If consistency holds when all the parameters are estimated simultaneously and when the number of examinees and the number of items becomes large simultaneously, then it seems reasonable to represent the information by:

$$(4) \qquad E\left[\frac{\delta \ln(L)}{\delta \xi_p}\,\frac{\delta \ln(L)}{\delta \xi_q}\right]$$

for $p,q = 1,2, \ldots, m$ and $m = 2n+N-2$. The likelihood is L and $\xi = \{\xi_p\} = \{\mu, \theta\}$, where

$\mu = \{a_1,b_1,c_1,a_2,b_2,c_2,\ldots,a_n,b_n,c_n\}$ and
$\theta = \{\theta_1, \theta_2, \ldots, \theta_{N-2}\}$

Lord and Wingersky (1985) suggested to gather the information in the response data about the m parameters in the following partitioned matrix:

$$
(5)\qquad \mathrm{Inf}_3 = \left[\begin{array}{ccccc|ccccc}
I_1 & & & & & K_{11} & K_{12} & \cdots & K_{1(N-2)} \\
 & I_2 & & 0 & & K_{21} & K_{22} & \cdots & K_{2(N-2)} \\
 & & \ddots & & & \vdots & & & \vdots \\
 & 0 & & & I_n & K_{n1} & K_{n2} & \cdots & K_{n(N-2)} \\
\hline
K'_{11} & K'_{21} & & \cdots & K'_{n1} & J_1 & & & \\
K'_{12} & K'_{22} & & \cdots & K'_{n2} & & J_2 & & 0 \\
\vdots & & & & \vdots & & & \ddots & \\
K'_{(N-2)1} & K'_{(N-2)2} & & \cdots & K'_{(N-2)n} & & 0 & & J_{(N-2)}
\end{array}\right] = \left[\begin{array}{c|c} I & K \\ \hline K' & J \end{array}\right]
$$

The 3nx3n super diagonal matrix I contains the 3x3 item information matrices I_1 through I_n for the parameters a_i, b_i, and c_i, respectively, for n different items. The (N-2)x(N-2) diagonal matrix J contains Fisher's information J_1 through J_{N-2} for the ability parameters θ_a and K_{ia} is the 3x1 joint Fisher information vector for item i and person a. Thus the information in the data for the estimation of the m parameters of the three-parameter logistic model simultaneously is stored in the information matrix Inf_3 and the asymptotic variances and covariances of the m estimated parameters can be obtained from $\mathrm{Cov}(\hat{\xi}) = \mathrm{Inf}_3^{-1}$.

Finally it should be noted that the information matrices Inf_2 and Inf_1 corresponding with the two- and one-parameter logistic model can be obtained from Inf_3 by deleting the appropriate rows and columns of I_1 through I_n and the corresponding elements of K_{ia}.

Two Seemingly Different Approaches to Estimate Standard Errors.

Two approaches have been employed to obtain the asymptotic standard errors of the estimated item parameters under the assumption that the examinees are not randomly sampled from a certain population. The difference between these two approaches is that the first approach (Thissen and Wainer, 1982; de Gruijter, 1984) assumes the ability parameters θ_a to be known, whereas the second approach (Lord and Wingersky, 1985; Wingersky and Lord, 1985; de Gruijter, 1988), does not assume the θ_a's to be known but takes into account the fact that they have to be estimated. In this paper it will be argued, that the approach that assumes the ability parameters θ_a to be unknown is in fact implemented by using the information on the item parameters corrected for the joint information with the ability parameters θ_a.

The first approach uses the fact that the maximum likelihood estimator $\hat{\mu}_i$ of the triplet $\mu_i = (a_i, b_i, c_i)$ is asymptotically normally distributed with covariance matrix $Cov(\hat{\mu}_i)$, which can be obtained from the inverse of the information matrix:

$$(6) \quad I = \begin{bmatrix} I_1 & & & 0 \\ & I_2 & & \\ & & \ddots & \\ 0 & & & I_n \end{bmatrix}$$

The elements of the 3x3 information matrices I_1 through I_n can be obtained from Lord (1980, p. 191) and the standard errors of the estimated parameters can be obtained from the diagonal elements of $Cov(\hat{\mu}) = I^{-1}$.

The second approach is based on the full information matrix Inf_3 (or Inf_2 and Inf_1 for the two- and one-parameter model) and assumes that the N ability parameters θ_a are unknown and have to be estimated jointly with the item parameters. The covariance matrix of the estimated item parameters is now denoted by $Cov^*(\hat{\mu})$ and is obtained from the leading 3nx3n matrix of Inf_3^{-1} (or Inf_2^{-1}, Inf_1^{-1}). The asymptotic standard errors of the estimated parameters are obtained from the diagonal elements of $Cov^*(\hat{\mu})$.

Information and 'Partial' Information

To demonstrate the argument that the two approaches which use the diagonal elements of $Cov(\hat{\mu})$ and $Cov^*(\hat{\mu})$ in fact both do not take into account the estimation of the ability parameters, the following result on partitioned matrices must be given.

Let H be an mxm non-singular tranformation matrix which is partitioned as $H = [H_1 \quad H_2]$. The mxp matrix H_1 and the

mx(m-p) matrix H_2 are of ranks $p \leq m$ and (m-p) respectively. The information matrix Inf_3 of m parameters can be transformed into a partitioned matrix containing information on two different sets of transformed parameters:

$$(7) \qquad H' Inf_3 H = \left[\begin{array}{c|c} H_1' Inf_3 H_1 & H_1' Inf_3 H_2 \\ \hline H_2' Inf_3 H_1 & H_2' Inf_3 H_2 \end{array}\right]$$

Let T be a pxm matrix, such that T H= [Id O]. Id is an identity matrix and O is a zero matrix. The columns of H_1 form the basis of the vector space spanned by the rows of T and H_2 is a so-called deficiency matrix. Then if the information matrix Inf_3 is symmetric positive definite, the following equality holds:

$$(8) \quad Inf_3^{-1} - Inf_3^{-1}T'(TInf_3^{-1}T')^{-1}TInf_3^{-1} = H_2(H_2'Inf_3H_2)^{-1}H_2'$$

(Lemma 1 of Khatri, 1966, p.76-77)

From (8) the information on the first set of transformed parameters given the information on the second set of transformed parameters can be obtained. Pre- and post-multiplication of both sides of (8) by $H_1' Inf_3$ and $Inf_3 H_1$, respectively, and rearrangement of terms yields

$$(9) \quad H_1'Inf_3H_1 - H_1'Inf_3H_2(H_2'Inf_3H_2)^{-1}H_2'Inf_3H_1 = [TInf_3^{-1}T']^{-1}.$$

Although H and T may be chosen arbitrarily, as long as the condition T H = [Id O] is satisfied, a very simple choice for H and T is:

$$(10) \quad [H_1 \; H_2] = \begin{bmatrix} Id & 0 \\ 0 & Id \end{bmatrix}, \quad T = H_1'.$$

(3n) (N-2)

This choice give the same partitioning of Inf_3 as given in (5) and (9) reduces to:

$$(11) \quad I - K\,(J)^{-1}K' = [Cov^*(\mu)\hat{}]^{-1}$$

This equation is comparable with equations 8 and 14 from de Gruijter (1988) for the one- and two-parameter model. From (11) it can be concluded that if we only consider the standard errors of the item parameters taken from the diagonal elements of $Cov^*(\hat{\mu})$, which is the 3nx3n leading matrix of the inverse of the full-information matrix Inf_3, we are in fact using the inverse of the 'partial' information on the item parameters corrected for their joint information with the ability parameters. So, the used information on the item parameters does not take into account the estimation of the ability parameters, i.e. the asymptotic covariances

between the estimates of the item parameters and the ability parameters are not considered.

An Optimality Criterion for Efficiency

From the previous section it is clear that there is nothing wrong in using the matrix $Cov^{*}(\hat{\mu})$ as an asymptotic efficiency measure. However, if the m parameters of an IRT model are all unknown, it would be better to use one of the measures common in optimal design research.

Suppose that we have a multivariate normal probability density of the estimator of a 1xm parameter vector ξ, given by:

$$(12) \quad p(\hat{\xi}) = (2\pi)^{-m/2} |Cov(\hat{\xi})|^{-1/2} \exp[-1/2 (\xi-\hat{\xi})' \{Cov(\hat{\xi})\}^{-1} (\xi-\hat{\xi})]$$

where $Cov(\hat{\xi})$ is the asymptotic variance-covariance matrix of $\hat{\xi}$. Shannon (1948) showed that a suitable measure of uncertainty associated with the probability density is

$$(13) \quad h(p(\hat{\xi})) = -E[\ln(p(\hat{\xi}))] \ .$$

Substitution of (12) into (13) yields

$$h(p(\hat{\xi})) = -E\left\{ {}_{1/2}[-m \ln(2\pi) - \ln|Cov(\hat{\xi})| - (\xi-\hat{\xi})' \{Cov(\hat{\xi})\}^{-1}(\xi-\hat{\xi})] \right\}$$

$$= {}_{1/2}\left\{ m \ln(2\pi) + \ln|Cov(\hat{\xi})| + Tr\ [\{Cov(\hat{\xi})\}^{-1} E(\xi-\hat{\xi})'(\xi-\hat{\xi})] \right\}$$

$$= {}_{1/2}\left\{ m\ (\ln(2\pi)+1) + \ln|Cov(\hat{\xi})| \right\}$$

Thus apart from the constant terms ${}_{1/2}$ and ${}_{1/2}m(\ln(2\pi)+1)$, $\ln|Cov(\hat{\xi})|$ is an adequate measure of uncertainty and can be obtained directly from $|Inf_3|^{-1}$. This criterion, the determinant of the covariance matrix of the estimated parameters, is often referred to as the generalized variance (Anderson, 1984), or D-optimality criterion.

It should be remarked, however, that while the generalized variance is a reasonable criterion to consider if one is interested in the whole set of m parameters in the model, this criterion is model dependent. If a design is investigated under the assumption that a model with m parameters is correct, but in fact a model with m-1 parameters is more appropriate, then it is impossible to compare the efficiencies of these two models with this criterion.

Finally, it should be noted that if the item parameters are the parameters of primary interest, then the marginal distribution of the estimator of the item parameters μ is needed. The marginal probability density of the estimator $\hat{\mu}$ of the item parameters (Graybill, 1969) is:

$$(14)\quad p(\hat{\mu}) = (2\pi)^{-3n/2}|Cov^*(\hat{\mu})|^{-1/2}\exp[-1/2(\mu-\hat{\mu})'\{Cov^*(\hat{\mu})\}^{-1}(\mu-\hat{\mu})]$$

where $Cov^*(\hat{\mu})$ is the leading 3nx3n matrix of $Cov(\hat{\xi})$. Following the same line of reasoning as above, a suitable measure of uncertainty would now be:

$$h^* = |Cov^*(\hat{\mu})|.$$

By means of the well-known determinant equation for partitioned matrices h^{*-1} can be computed from:

$$(15)\quad h^{*-1} = |\, I - K\,(J)^{-1}K'\,|$$

Item Informations for Different IRT Models

Thissen and Wainer (1982) compared the asymptotic standard errors for a_i, b_i and c_i in the one-, two-, and three parameter logistic models, under the assumption that the ability parameters are all known. For easy items, a huge standard error for b_i was found. Thissen and Wainer (1982) explained this result by referring to the large standard error for the estimate of the lower asymptote c_i for easy items and the large covariance between the lower asymptote c_i and the location parameter b_i, causing the uncertainty about the lower asymptote for easy items to move from the estimate of c_i to the estimate of b_i. An alternative explanation, however, based upon the concept of 'partial' information can be given.

Let I_i (i=1,...,n) be the 3x3 item information matrix for the three-parameter logistic model:

$$I_i = \begin{bmatrix} I_{aa} & & \text{symm.} \\ I_{ab} & I_{bb} & \\ I_{ac} & I_{bc} & I_{cc} \end{bmatrix}$$

and let Cov_{bb} be the asymptotic variance of the estimator for parameter b_i. For each of the 3 IRT models Cov_{bb} can be written in terms of the information connected with each parameter as follows:

$$Cov_{bb} = I_{bb}^{-1}$$

for the one-parameter model,

$$Cov_{bb} = (I_{bb} - I_{ab} (I_{aa})^{-1} I_{ab})^{-1}$$

for the two-parameter model and

$$Cov_{bb} = \{I_{bb} - [I_{ab} \; I_{bc}] \begin{bmatrix} I_{aa} & I_{ac} \\ I_{ac} & I_{cc} \end{bmatrix}^{-1} \begin{bmatrix} I_{ab} \\ I_{bc} \end{bmatrix}\}^{-1}$$

for the three-parameter model.

The difference between the variances of b_i for these three models is in fact a difference between the information on b_i and the 'partial' information on b_i, corrected for the joint information on b_i and a_i and/or c_i. In other words, the asymptotic variance Cov_{bb} in the one-parameter model is equal to the asymptotic 'partial' variance of the estimator b_i

given an estimate for a_i and/or c_i in the three-parameter model.

From the formula $I_{cc} = \{\sum_a [1-P_i(\theta_a)]P_i(\theta_a)\}/(1-c_i)$ (Lord, 1980, p. 191) it can be concluded that the information I_{cc} on the lower asymptote c_i decreases rapidly as the difficulty parameter b_i decreases and becomes relatively small for very easy items because there are very few observations to estimate c_i when an item is easy (Thissen and Wainer, 1982). This means that for small values of b_i the elements of

$$\begin{bmatrix} I_{aa} & I_{ac} \\ I_{ac} & I_{cc} \end{bmatrix}^{-1}$$

will become relatively large, which means that for easy items the 'partial' information of b_i will be much smaller than the information of b_i.

Comparison of Designs

Suppose that an instructional programme is designed to teach a number of skills, which are grouped in a domain of proficiency skills, to a population of examinees. Assume also that an item domain can be distinguished containing a set of items that may be used to assess mastery of each individual skill and that these items can be ranked according to their difficulties.

A design $D(\mu)$ can be characterized by the sampling procedure used to select examinees from the population and by the selection of items from the item domain. For the one-, two- and three-parameter model the parameter set is $\mu_i=\{b_i\}$, $\mu_i=\{a_i,b_i\}$, $\mu_i=\{a_i,b_i,c_i\}$, respectively.

The relative efficiency of a particular design $D_1(\mu)$ related to another design $D_2(\mu)$, both used to estimate the same parameter set μ_i is:

$$(16) \qquad RE\{D_1(\mu),D_2(\mu)\} = \frac{|\,Cov_2(\hat{\mu}_i)\,|}{|\,Cov_1(\hat{\mu}_i)\,|}$$

where $|Cov_1(\hat{\mu}_i)|$ and $|Cov_2(\hat{\mu}_i)|$ are the generalized variances of the estimated parameters in $D_1(\mu)$ and $D_2(\mu)$, respectively. If the relative efficiency is less than one, the design $D_1(\mu)$ is less efficient than $D_2(\mu)$, and if the relative efficiency is greater than one, the converse is true.

It is well known that for fixed values of the discrimination parameter a_i and for $c_i=0$ the information $I_{bb}= a_i\{\Sigma P_i(\theta_a)[1-P_i(\theta_a)]\}$ is maximal for $P_i(\theta_a)=0.5$, i.e. when $\theta_a= b_i$. Thus, a sample of examinees with abilities all equal to the item difficulties would give optimal information if only the difficulty parameter is estimated (van der Linden, 1988). If, however, the two- and three parameter models are considered this will not generally be true.

Insert Figure 1 about here

In Figure 1 the relative efficiency for the one-, two- and three-parameter model is given as a function of the standard deviation of the abilities. This relative efficiency is computed by relating the $Cov(\hat{\mu}_i)$ of designs where examinees are sampled from normal populations with standard deviation of ability $SD_\theta \in \langle 2 , 0.1 \rangle$ to that of a design where examinees come from a normal population with $SD_\theta = 2.0$. Decreasing values of the SD_θ result in an increasing relative efficiency for the one-parameter model. For the two-parameter model, however, decrease of standard deviation of ability will not generally lead to increase of efficiency. The relative efficiency first increases and eventually decreases as the standard deviation of ability becomes smaller. This effect can be explained by the fact that the variance of the estimated discrimination (slope) parameter generally increases as the standard deviation of ability becomes smaller. A similar effect can be found in regression analysis, where the variance of the estimated slope is inversely related to the variance of the independent variable.

The pattern of relative efficiencies for the three-parameter model is reverse to that of the one-parameter model. The most efficient design for a three-parameter model will be one in which the abilities have a relative large

standard deviation. This result is mainly caused by the problems connected with the estimation of the lower asymptote c_i for easy items.

The relative efficiencies in Figure 1 suggest that for the one- and two-parameter model efficiency will be gained by sampling examinees from subpopulations with relative small differences in ability among examinees. For a test of items with difficulties $b_i \in \langle -3,+3 \rangle$ it would be more efficient, for example, to select two samples of examinees with abilities $\theta_a \in \langle -3, 0 \rangle$ and $\theta_a \in \langle 0,+3 \rangle$, respectively, and administer items with corresponding difficulties $b_i \in \langle -3,0 \rangle$ and $b_i \in \langle 0,+3 \rangle$ to each of these samples, than to administer items with $b_i \in \langle -3,+3 \rangle$ to one sample of examinees with $\theta_a \in \langle -3,+3 \rangle$.

Insert Figure 2 about here

In Figure 2 seven different designs are diagrammed for a fixed number of examinee-item confrontations and depending on the range of the ability scale some items are assumed to be anchor items in the designs $D_2(\mu)$, $D_4(\mu)$, $D_5(\mu)$ and $D_6(\mu)$.

Insert Figures 3, 4 and 5 about here

In the Figures 3,4 and 5 the relative efficiencies of $D_2(\mu)$ through $D_7(\mu)$ compared to design $D_1(\mu)$ are given. It is clear from these figures that for the one- and two-parameter model design $D_3(\mu)$ is more efficient than $D_2(\mu)$, that design $D_5(\mu)$ is more efficient than $D_4(\mu)$, and that $D_7(\mu)$ is more efficient than $D_6(\mu)$. This result can be explained by the smaller standard deviations of ability in $D_3(\mu)$, $D_5(\mu)$ and $D_7(\mu)$ and is also in agreement with the conclusion drawn by Vale (1986), that the number of anchor items does not improve efficiency.

For the one-parameter model design $D_7(\mu)$ seems the most efficient one. For the two-parameter model the pattern is not consistent. Design $D_3(\mu)$ seems very efficient, except for items with difficulty $b_i = 0$. This is explained by the fact that for $b_i=0$ the discrimination parameter cannot be estimated very efficiently in $D_3(\mu)$. Design $D_7(\mu)$, however, seems to be a good alternative to $D_3(\mu)$, especially for $b_i=0$ and large values of a_i.

For the three-parameter model designs $D_6(\mu)$ and $D_7(\mu)$ are most efficient for very easy items, but for $b_i > 0$ design $D_1(\mu)$ is superior.

Discussion

As the sampling concept from survey analysis gained currency in educational measurement, the interest in alternative sampling designs increased. Especially in large scale assessment studies it would save a lot of classroom

administration time if only a set of items instead of the whole test would be administered to the same examinees. The answer to the question whether such a selection of items would influence the efficiency of the IRT parameter estimates will depend on the assumed IRT model.

Suppose that a test which measures certain skills is considered and suppose that pupils can be grouped according to their ability to master such skills. If it is also possible to administer the easy items from that test to the less skilled pupils and the difficult items to the more skilled pupils, then the following recommendations for efficient estimation of the item parameters can be given:

- For the one- and two-parameter model it would be more efficient to administer the easy part of a test to a sample of N less skilled pupils and the difficult part to another sample of N more skilled pupils, than to administer the whole test to a sample of size N, containing both less and more skilled pupils. This would also be the case for the three-parameter model when easy highly discriminating items are estimated. Moreover, a lot of classroom administration time would be saved (Design D_1 versus D_3).
- For all three IRT models it would be generally better to administer the whole test to a sample of N pupils than to administer about 70% of the test to two different samples of 0.7N pupils, such that the easy part of the test is administered to the less skilled group and the difficult part to the more skilled group (Design D_1 versus D_2).

- Although anchor items are primarily used to establish a common scale for the item or the ability parameters, the use of anchor items will generally not increase efficiency. This result supports the conclusion drawn by Vale (1986).
- For the efficient estimation of parameters of the three-parameter model it is generally more efficient to use one sample of pupils with a relatively large SD_θ. Only for easy items efficiency is gained by administering very easy items to a separate sample of less skilled pupils.

References

Anderson, T.W. (1984). An introduction to multivariate statistical analysis, 2nd ed., New York: Wiley.

de Gruijter, D.N.M. (1984). A comment on some standard errors in item response theory. Psychometrika, 49, 269-272.

de Gruijter, D.N.M. (1985). A note on the asymptotic variance-covariance matrix of item parameter estimates in the Rasch model. Psychometrika, 50, 247-249.

de Gruijter, D.N.M. (1988). Standard errors of item parameter estimates in incomplete designs. Applied Psychological Measurement, 12, 109-116.

Graybill, F.A. (1969) Introduction to matrices with applications in statistics, California Belmont: Wadsworth Publ. Comp.

Hambleton, R.K. & Swaminathan, H. (1985). Item response theory, Boston: Kluwer-Nijhoff Publishing Comp.

Kendall, M.G. & Stuart, A. (1973). The advanced theory of statistics (Vol 2), New York: Hafner.

Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curve. Annals of the Institute of Statistical Mathematics, 18, 75-86.

Lord, F.M. (1962). Estimating norms by item-sampling. Educational and Psychological Measurement, 22, 259-267.

Lord, F.M. (1974). The relative efficiency of two tests as a function of ability, Psychometrika, 39, 351-358.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale N.J.: Erlbaum Associates.

Lord, F.M. & Wingersky, M.S. (1985). Sampling variances and covariances of parameter estimates in item response theory. In D.J. Weiss (ed.), Proceedings of the 1982 Item Response Theory and Computerized Adaptive Testing Conference. pp. 69-88, Minneapolis: University of Minnesota.

Samejima, F. (1977). A use of the information function in tailored testing. Applied Psychological Measurement, 1, 233-247.

Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423, 623-656.

Theunissen, T.J.J.M. (1985). Binary programming and test design. Psychometrika, 50, 411-420.

Thissen, D. & Wainer, H. (1982). Some standard errors in item response theory. Psychometrika, 47, 397-412.

van der Linden, W.J. (1987). IRT-based test construction (Research Report 87-2) Enschede: University of Twente.

van der Linden, W.J. (1988). Optimizing incomplete sampling designs for item response model parameters (Research Report 88-5) Enschede: University of Twente.

van der Linden, W.J. & Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. Psychometrika. in press.

Vale, C.D. (1986). Linking item parameters onto a common scale. Applied Psychological Measurement, 10, 333-344.

Wingersky, M.S. & Lord, F.M. (1985). An investigation of methods for reducing sampling error in certain IRT procedures. Applied Psychological Measurement, 8, 347-364.

Figure captions

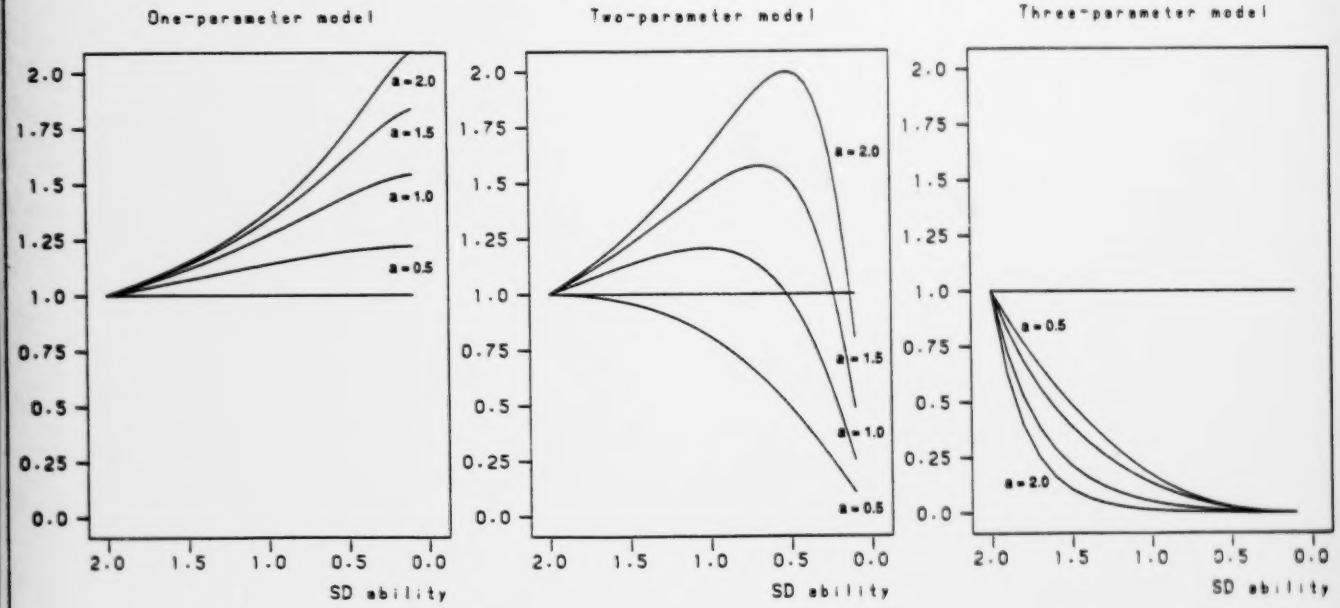
Figure 1. Relative efficiency of designs with standard deviations for ability $2.0 \leq SD_\theta \leq 0.2$

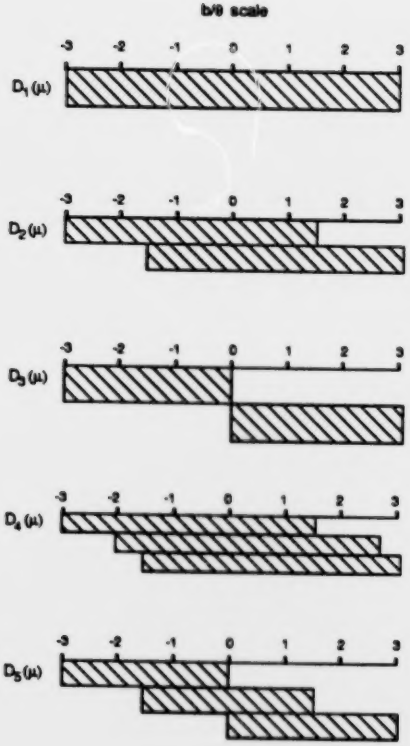
Figure 2. Seven designs for a fixed number of examinee-item confrontations

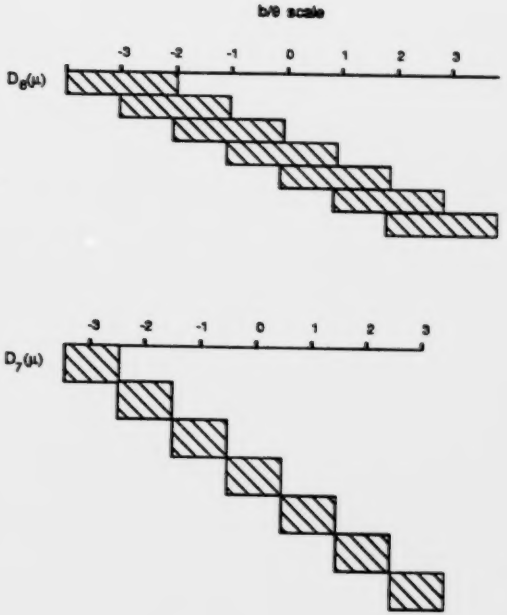
Figure 3. Relative efficiency of designs $D_2(\mu)$ and $D_3(\mu)$ related to $D_1(\mu)$

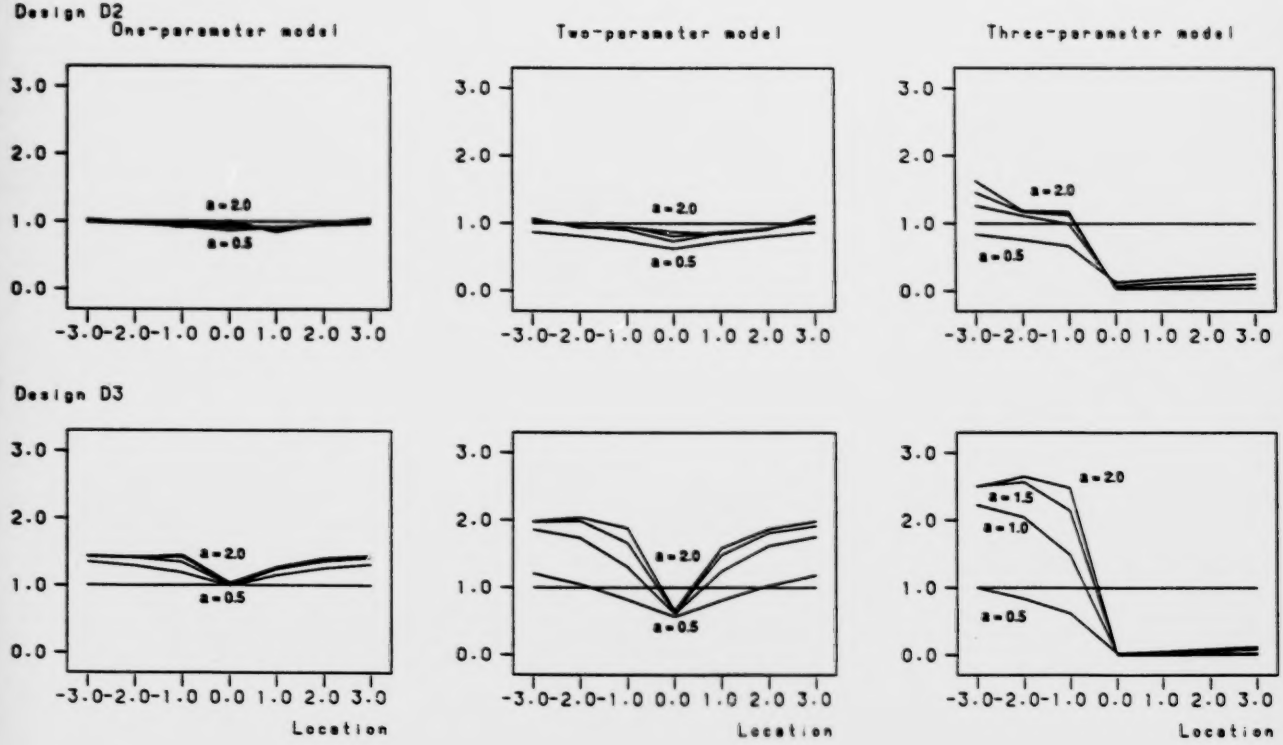
Figure 4. Relative efficiency of designs $D_4(\mu)$ and $D_5(\mu)$ related to $D_1(\mu)$

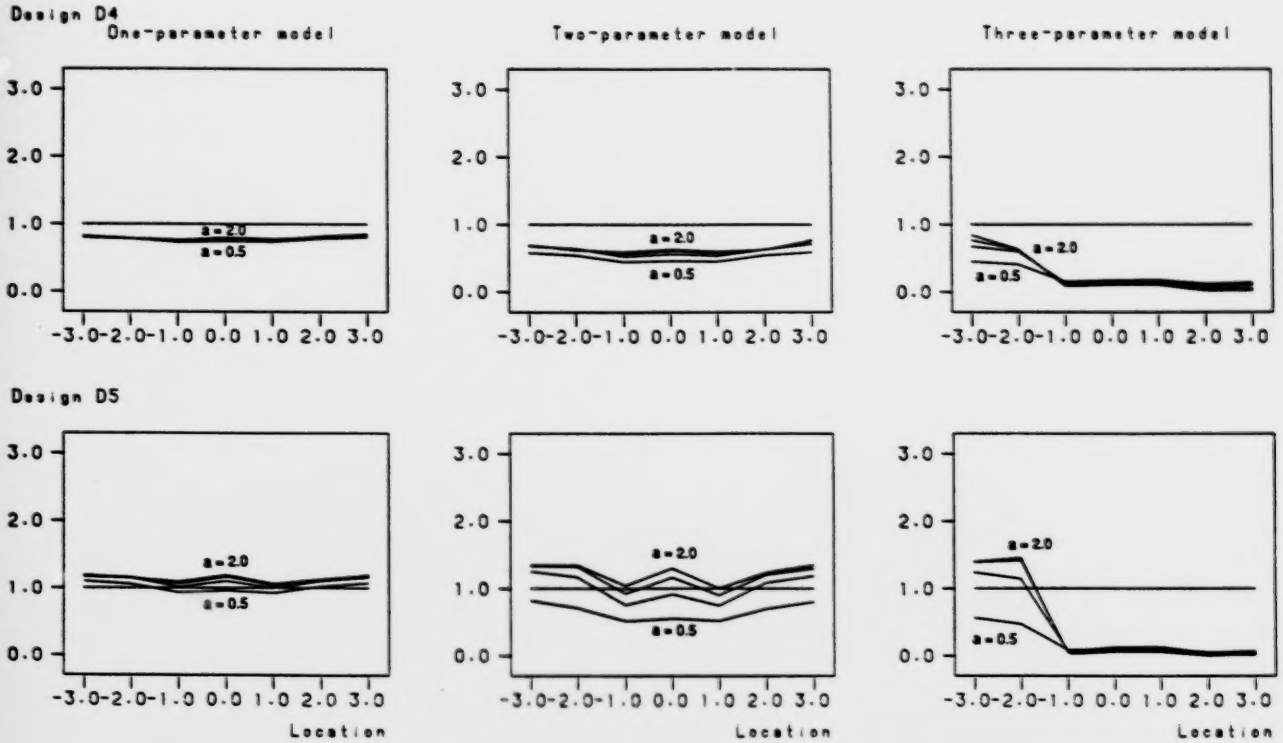
Figure 5. Relative efficiency of designs $D_6(\mu)$ and $D_7(\mu)$ related to $D_1(\mu)$


One-parameter model
Two-parameter model
Three-parameter model
2.0
1.75
1.5
1.25
1.0
0.75
0.5
0.25
0.0
a=2.0
a=1.5
a=1.0
a=0.5
2.0 1.5 1.0 0.5 0.0
SD ability

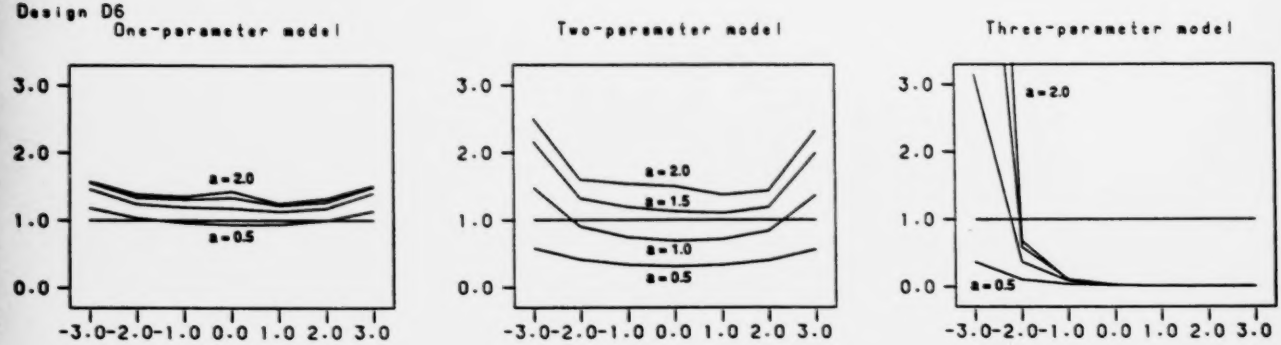

b/θ scale
-3 -2 -1 0 1 2 3
$D_1(\mu)$
$D_2(\mu)$
$D_3(\mu)$
$D_4(\mu)$
$D_5(\mu)$


b/θ scale
-3 -2 -1 0 1 2 3
$D_6(\mu)$
$D_7(\mu)$

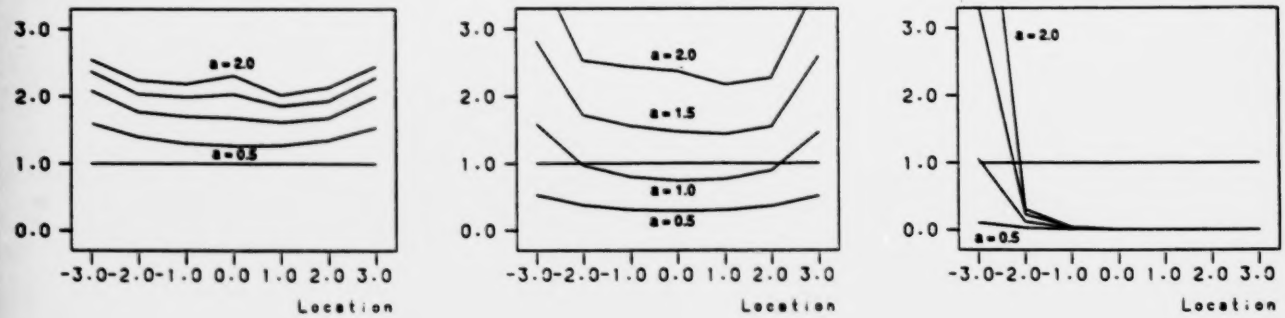

Design D2
One-parameter model
Two-parameter model
Three-parameter model
Design D3
a=2.0
a=1.5
a=1.0
a=0.5
3.0
2.0
1.0
0.0
-3.0 -2.0 -1.0 0.0 1.0 2.0 3.0
Location


Design D4
One-parameter model
Two-parameter model
Three-parameter model
3.0
2.0
1.0
0.0
a=2.0
a=0.5
-3.0 -2.0 -1.0 0.0 1.0 2.0 3.0
Design D5
Location

Design D6



Design D7



Titles of recent Research Reports from the Division of Educational Measurement and Data Analysis, University of Twente, Enschede, The Netherlands.

RR-88-1 E. van der Burg & J. de Leeuw, *Nonlinear redundancy analysis*

RR-88-2 W.J. van der Linden & J.J. Adema, *Algorithmic test design using classical item parameters*

RR-88-3 E. Boekkooi-Timminga, *A cluster-based method for test construction*

RR-88-4 J.J. Adema, *A note on solving large-scale zero-one programming problems*

RR-88-5 W.J. van der Linden, *Optimizing incomplete sample designs for item response model parameters*

RR-88-6 H.J. Vos, *The use of decision theory in the Minnesota Adaptive Instructional System*

RR-88-7 J.H.A.N. Rikers, *Towards an authoring system for item construction*

RR-88-8 R.J.H. Engelen, W.J. van der Linden, & S.J. Oosterloo, *Item information in the Rasch model*

RR-88-9 W.J. van der Linden & T.J.H.M. Eggen, *The Rasch model as a model for paired comparisons with an individual tie parameter*

RR-88-10 H. Kelderman & G. Macready, *Loglinear-latent-class models for detecting item bias*

RR-88-11 D.L. Knol & M.P.F. Berger, *Empirical comparison between factor analysis and item response models*

RR-88-12 E. van der Burg & G. Dijksterhuis, *Nonlinear canonical correlation analysis of multiway data*

RR-88-13 J. Kogut, *Asymptotic distribution of an IRT person fit index*

RR-88-14 J.J. Adema, *The construction of two-stage tests*

RR-88-15 H.J. Vos, *Simultaneous optimization of decisions using a linear utility function*

RR-88-16 H. Kelderman, *An IRT model for item responses that are subject to omission and/or intrusion errors*

RR-88-17 H. Kelderman, *Loglinear multidimensional IRT models for polytomously scored items*

RR-88-18 H.J. Vos, *Applications of decision theory to computer based adaptive instructional systems*

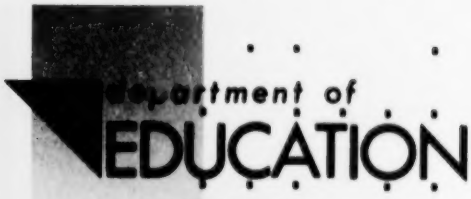
RR-89-1 R.J.H. Engelen & R.J. Jannarone, *A connection between item/subtest regression and the Rasch model*

RR-89-2 E. Boekkooi-Timminga, *The construction of parallel tests from IRT-based item banks*

RR-89-3 D.L. Knol, *Stepwise item selection procedures for Rasch scales using quasi-loglinear models*

RR-89-4 M.P.F. Berger, *On the efficiency of IRT models when applied to different sampling designs*

Research Reports can be obtained at costs from Bibliotheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.



A publication by
the Department of Education
of the University of Twente
P.O. Box 217
ede
nds